

CLAIMS

What is claimed is:

1. In a computer system, a method for clustering a plurality of datapoints, wherein each datapoint is a series of gene expression values, wherein the method comprises.
- 5 Sub CS
- a) receiving the gene expression values of the datapoints;
- b) using a self organizing map, clustering the datapoints such that the datapoints that exhibit similar patterns are clustered together into respective clusters; and
- 10 c) providing an output indicating the clusters of the datapoints.
2. The method of Claim 1, wherein the gene expression values are obtained from a gene that is subjected to at least one condition.
3. The method of Claim 2, the step of receiving includes receiving gene expression values of datasets, wherein a dataset is a series of gene expression values across multiple genes for a condition.
- 15
4. The method of Claim 3, further comprising filtering out any datapoints that exhibit an insignificant change in the gene expression value, such that working datapoints remain.
5. The method of Claim 4, further comprising normalizing the gene expression value of the working datapoints.
- 20

004460-04400

6. The method of Claim 5, wherein the self organizing map is formed of a plurality of Nodes, N, and clusters the datapoints according to a competitive learning routine.

7. The method of Claim 6, wherein the competitive learning routine is:

$$f_{i+1}(N) = f_i(N) + \tau(d(N, N_p), i) (P - f_i(N))$$

wherein i = number of iterations, N = the node of the self organizing map, τ = learning rate, P = the subject working datapoint, d = distance, N_p = node that is mapped nearest to P , and $f_i(N)$ is the position of N at i .

8. The method of Claim 1, wherein the step of providing includes displaying at least one representative datapoint from each cluster.

9. The method of Claim 5, wherein the step of normalizing the gene expression value comprises determining the ratio of a) difference between the subject gene expression value and the average gene expression value across datasets, and b) the standard deviation of the gene expression value across datasets.

10. The method of Claim 3, further comprising rescaling the gene expression values to account for variations across multiple conditions.

11. In a computer system, a method for grouping a plurality of datapoints, wherein each datapoint is a series of gene expression values, wherein the method comprises:

- a) receiving gene expression values of the datapoints;
- b) filtering out any datapoints that exhibit an insignificant change in the gene expression value, such that working datapoints remain;
- c) normalizing the gene expression value of the working datapoints;

- d) using a self organizing map, grouping the working datapoints such that the datapoints that exhibit similar patterns are grouped together into respective clusters; and
- e) providing an output indicating the groups of the datapoints.

5 12. The method of Claim 11, wherein the gene expression values are obtained from a gene that is subjected to at least one condition.

13. The method of Claim 12, the step of receiving includes receiving gene expression values of datasets, wherein a dataset is a series of gene expression values across multiple genes for a condition.

10 14. The method of Claim 13, wherein the self organizing map is formed of a plurality of Nodes, N, and groups the datapoints according to a competitive learning routine.

15. The method of Claim 14, wherein the competitive learning routine is:

$$f_{i+1}(N) = f_i(N) + \tau(d(N, N_p), i) (P - f_i(N))$$

15 wherein i = number of iterations, N= the node of the self organizing map, τ = learning rate, P = the subject working datapoint, d = distance, N_p = node that is mapped nearest to P, and $f_i(N)$ is the position of N at i.

16. The method of Claim 11, wherein the step of providing includes displaying at least one representative datapoint from each group.

20 17. The method of Claim 13, wherein the step of normalizing the gene expression value comprises determining the ratio of a) difference between the subject gene

expression value and the average gene expression value across datasets, and b) the standard deviation of the gene expression value across datasets.

18. The method of Claim 11, further comprising rescaling the gene expression values to account for variations across multiple conditions.
- 5 19. A computer apparatus for clustering a plurality of datapoints, wherein each datapoint is a series of gene expression values, wherein the apparatus comprises:
- a) a source of gene expression values of the datapoints;
 - b) a processor routine coupled to receive datapoints from the source, the processor routine utilizing a self organizing map for clustering datapoints such that the datapoints that exhibit similar patterns are clustered together into respective clusters; and
 - 10 c) an output device, coupled to the processor routine, for indicating the clusters of the datapoints.
20. The apparatus of Claim 19, wherein the gene expression values are obtained from a gene that is subjected to at least one condition.
- 15 21. The apparatus of Claim 20, wherein the source further provides datasets, each dataset is a series of gene expression values across multiple genes for a condition.
22. The computer apparatus of Claim 21, further comprising a filter, coupled to the source, for filtering out any of the datapoints that exhibit an insignificant change in the gene expression value, such that working datapoints remain.
- 20

23. The computer apparatus of Claim 22, further comprising a normalizing processor coupled to the filter, for normalizing the gene expression value of the working datapoints.
24. The computer apparatus of Claim 23, wherein the normalizing process determines a normalized gene expression value according to the ratio of a) difference between the subject gene expression value and the average gene expression value across datasets, and b) the standard deviation of the gene expression value across datasets.
25. The computer apparatus of Claim 24, wherein the self organizing map is formed of a plurality of Nodes, N, and clusters the datapoints according to a competitive learning routine.
26. The computer apparatus of Claim 25, wherein the competitive learning routine is:
- $$f_{i+1}(N) = f_i(N) + \tau(d(N, N_p), i) (P - f_i(N))$$
- wherein i = number of iterations, N = the node of the self organizing map, τ = learning rate, P = the subject working datapoint, d = distance, N_p = node that is mapped nearest to P, and $f_i(N)$ is the position of N at i.
27. The computer apparatus of Claim 26, wherein the output device comprises a display of at least one representative datapoint from each cluster.
28. A computer apparatus for grouping a plurality of datapoints, wherein each datapoint is a series of gene expression values, wherein the apparatus comprises:
- a source of gene expression values of the datapoints;

- 5
- b) a filter, coupled to the source, for receiving the gene expression values and filtering out any of the datapoints that exhibit an insignificant change in the gene expression value, such that working datapoints remain;
 - c) a normalizing process, coupled to the filter, for normalizing the gene expression value of the working datapoints;
 - d) a processor routine that is responsive to the normalizing process and utilizes a self organizing map for grouping the working datapoints such that the datapoints that exhibit similar patterns are grouped together into respective groups; and
 - 10 e) an output device, coupled to the processor routine, for indicating the groups of the datapoints.

29. The apparatus of Claim 28, wherein the gene expression values are obtained from a gene that is subjected to at least one condition.

15 30. The apparatus of Claim 29, wherein the source further provides datasets, each dataset being a series of gene expression values across multiple genes for a condition.

31. The computer apparatus of Claim 22, wherein the normalizing process of the gene expression value is determined according to the ratio of a) difference between the subject gene expression value and the average gene expression value across datasets, and b) the standard deviation of the gene expression value across datasets.

20

32. The computer apparatus of Claim 31, wherein the self organizing map is formed of a plurality of Nodes, N, and groups the datapoints according to a competitive learning routine.

0047E0-2473250

33. The computer apparatus of Claim 32, wherein the competitive learning routine is:

$$f_{i+1}(N) = f_i(N) + \tau(d(N, N_p), i) (P - f_i(N))$$

wherein i = number of iterations, N = the node of the self organizing map, τ = learning rate, P = the subject working datapoint, d = distance, N_p = node that is mapped nearest to P , and $f_i(N)$ is the position of N at i .

34. The computer apparatus of Claim 33, wherein the output device comprises a display of at least one representative datapoint from each group.

35. A method for assessing expression patterns of two or more genes in cells, wherein the expression patterns are represented by a plurality of datapoints, wherein each datapoint is a series of gene expression values, wherein the method comprises:

- a) receiving the gene expression values of the datapoints;
- b) using a self organizing map, clustering the datapoints such that the datapoints that exhibit similar patterns are clustered together into respective clusters;
- c) providing an output indicating the clusters of the datapoints; and
- d) analyzing the output to determine the similarities or differences between the expression patterns of the genes.

36. The method of Claim 35, wherein the gene expression values are obtained from a gene that is subjected to at least one condition.

37. The method of Claim 36, wherein a dataset is a series of gene expression values across multiple genes for a condition.

38. The method of Claim 37, further comprising filtering out any datapoints that exhibit an insignificant change in the gene expression value, such that working datapoints remain.
39. The method of Claim 38, further comprising normalizing the gene expression value of the working datapoints.
40. The method of Claim 39, wherein the self organizing map is formed of a plurality of Nodes, N , and clusters the datapoints according to a competitive learning routine.
41. The method of Claim 40, wherein the competitive learning routine is:
- $$f_{i+1}(N) = f_i(N) + \tau(d(N, N_p), i) (P - f_i(N))$$
- wherein i = number of iterations, N = the node of the self organizing map, τ = learning rate, P = the subject working datapoint, d = distance, N_p = node that is mapped nearest to P , and $f_i(N)$ is the position of N at i .
42. The method of Claim 39, wherein the step of normalizing the gene expression value comprises determining the ratio of a) difference between the subject gene expression value and the average gene expression value across the datasets, and b) the standard deviation of the gene expression value across datasets.
43. The method of Claim 35, further comprising rescaling the gene expression values to account for variations across multiple conditions.
44. A method for characterizing expression patterns of a plurality of genes of a sample having unknown characteristics, wherein the sample from an individual is obtained and subjected to a multiplicity of diagnostic tests, and the expression

patterns of the genes for the diagnostic tests are represented by a plurality of datapoints, wherein the datapoint is a series of gene expression values across multiple genes for the diagnostic test, wherein the method comprises:

- a) receiving the gene expression values of the datapoints from the diagnostic tests;
 - b) using a self organizing map, clustering the datapoints such that the datapoints that exhibit similar patterns are clustered together into respective clusters;
 - c) providing an output indicating the clusters of the datapoints; and
 - d) comparing the output of the gene expression patterns of the unknown sample against a control,
- thereby characterizing gene expression patterns of the sample.

45. The method of Claim 44, wherein the gene expression values across multiple genes for the diagnostic test is obtained from a gene subjected to at least one condition.

46. The method of Claim 45, wherein a dataset is a series of gene expression values from a gene subjected to the diagnostic tests.

47. The method of Claim 46, wherein the sample from the individual is selected from the group consisting of: cells, lysed cells, cellular material suitable for determining gene expression, and material containing gene expression products.

48. The method of Claim 47, further comprising normalizing the gene expression value of the datapoints.

49. The method of Claim 48, wherein the self organizing map is formed of a plurality of Nodes, N, and clusters the datapoints according to a competitive learning routine.

50. The method of Claim 49, wherein the competitive learning routine is:

$$f_{i+1}(N) = f_i(N) + \tau(d(N, N_p), i) (P - f_i(N))$$

wherein i = number of iterations, N = the node of the self organizing map, τ = learning rate, P = the subject working datapoint, d = distance, N_p = node that is mapped nearest to P, and $f_i(N)$ is the position of N at i.

51. The method of Claim 50, wherein the step of normalizing the gene expression value comprises determining the ratio of a) difference between the subject gene expression value and the average gene expression value across datasets, and b) the standard deviation of the gene expression value across datasets.

52. A method of determining relatedness of expression patterns of two or more genes, wherein the expression patterns are represented by a plurality of datapoints, wherein each datapoint is a series of gene expression values, wherein the method comprises:

- a) receiving the gene expression values of the datapoints;
- b) using a self organizing map, clustering the datapoints such that the datapoints that exhibit similar patterns are clustered together into respective clusters;
- c) providing an output indicating the clusters of the datapoints; and
- d) analyzing the output to determine the similarities and/or differences between the expression patterns of the genes, thereby determining the relatedness of two or more genes.

004750-21152560

53. The method of Claim 52, wherein the gene expression values are obtained from a gene that is subjected to at least one condition.

54. The method of Claim 53, wherein a dataset is a series of gene expression values across multiple genes for a condition.

5 55. The method of Claim 54, further comprising filtering out any datapoints that exhibit an insignificant change in the gene expression value, such that working datapoints remain.

56. The method of Claim 55, further comprising normalizing the gene expression value of the working datapoints.

10 57. The method of Claim 56, wherein the self organizing map clusters the datapoints according to:

$$f_{i+1}(N) = f_i(N) + \tau(d(N, N_p), i) (P - f_i(N))$$

15 wherein i = number of iterations, N = the node of the self organizing map, τ = learning rate, P = the subject working datapoint, d = distance, N_p = node that is mapped nearest to P , and $f_i(N)$ is the position of N at i .

58. A method of identifying a drug target from the expression patterns of two or more genes from cells, the expression patterns are represented by a plurality of datapoints, and wherein each datapoint is a series of gene expression values, wherein the method comprises:

- 20
- a) obtaining cells that express genes,
 - b) subjecting the cells to an agent or condition for testing the drug target,
 - c) measuring gene expression from the cells subjected to the agent or condition, and from a control, to obtain the gene expression values,

- 5

10

60.

15

$$f_{i+1}(N) = f_i(N) + \tau(d(N, N_p), i) (P - f_i(N))$$

wherein i = number of iterations, N = the node of the self organizing map, τ = learning rate, P = the subject working datapoint, d = distance, N_p = node that is mapped nearest to P , and $f_i(N)$ is the position of N at i .